

Denison University

Denison Digital Commons

Student Scholarship

2024

A comparative Analysis of New York Times' Vietnam War Coverage: Pre-and Post-Pentagon Papers' Release

Khanh Do

Follow this and additional works at: <https://digitalcommons.denison.edu/studentscholarship>

A Comparative Analysis of the Evolution of New York Times' Vietnam War Coverage:

Pre- and Post-Pentagon Papers' Release

Khanh Do

Project Advisor: Professor Frank Proctor

Department of History – Digital Humanities Program

Denison University Summer Scholars

2024

Abstract

This study investigates how The New York Times' reporting of the Vietnam War transformed, particularly following the 1971 release of the Pentagon Papers using distant reading, a Digital Humanities technique that uses computers to read and analyze large bodies of texts. As a significant moment in media-government relations and a catalyst for changing public opinion and policy, the Pentagon Papers' impact is central to our analysis. Using the New York Times' comprehensive article archive, we trace the media's narrative shifts pre- and post-leak. Our methodology combines Digital Humanities and Computer Science, employing word cloud for theme identification¹, VADER for sentiment analysis², and Python for data processing and visualization. We aim to reveal changes in topics, sentiments, and linguistic styles in the newspaper's coverage before and after the Pentagon Papers event. The culmination of this research will be a publicly accessible website, serving as a repository for our findings, data visualizations, and a curated collection of these impactful news articles. This platform aims not only to shed light on the historical media portrayal of the Vietnam War but also to serve as a valuable resource for scholars, educators, and the general public interested in understanding the role of media in shaping historical narratives.

¹ Natural Language Processing technique that identifies common words in a large body of text

² Natural Language Processing technique that identifies the tone or emotion of words and phrases in a body of text by assigning numerical scores

Purpose of Study

This research aims to analyze the transformation in the New York Times' coverage of the Vietnam War, focusing on the role of the 1971 Pentagon Papers' release. Pentagon Papers, leaked by military analyst Daniel Ellsberg, exposed that the U.S. government had systematically lied to the public about the war's progress and the prospects for victory. The publication of the documents by The New York Times (NYT) profoundly impacted public opinion and how the media covered the war. The core objective of this study is to scrutinize the evolution of topics and sentiments in The New York Times' coverage.

The Vietnam War has been considered the first television war as some argued that media and television played a crucial role in the U.S. defeat. Early in the conflict, U.S. media, influenced by government-managed information, gradually transitioned to more independent journalism, relying on research, interviews, and analytical essays. In 1971, the Pentagon Papers articles on the New York Times's front page, detailing the Vietnam War's true purpose to the American people, forever changed the relationship between the people and the governing institution, completing the American people's disillusionment and distrust of the government. While technology has enabled historians and researchers to analyze historical events and narratives, including the Vietnam War, there remains a notable gap in data-driven studies examining how the New York Times' coverage topics and sentiments evolved pre- and post-Pentagon Papers. This research intends to fill this gap using advanced digital humanities techniques, combining my computer science and digital humanities background.

The project's overall goal was to comprehensively analyze the rhetoric, themes, and historical context embedded in these New York Times articles. In particular, we wanted to address these following research questions:

- How did the sentiment of media coverage regarding the Vietnam War change before and after the release of the Pentagon Papers?
- What were the predominant themes associated with political figures, geographical locations, and the human cost of the war during this period?

More than an academic endeavor, this project aims to deepen understanding of media responsibility, bias, and influence on public opinion in national crises. An accessible web platform will present these insights, bridging academic research and public engagement, thus providing an interactive, educational window into a critical historical period. Furthermore, as a Vietnamese, this research holds personal significance, offering a unique opportunity to explore American perspectives on the Vietnam War and their connection to Vietnam's history.

Literature Review

To analyze the difference in New York Times coverage during the Vietnam War pre- and post-Pentagon Papers, I first researched the background and timeline of the conflict and television war. Initially, U.S. media, including The New York Times, exhibited general support for American involvement in Vietnam. This support can be traced back to the early 1960s, a period marked by significant changes in U.S. and South Vietnamese leadership and escalating conflict. Media portrayal during this phase was often uncritical and supportive of the U.S. government's policies, reflecting Cold War ideologies and ethnocentric perspectives. However, as the war progressed, there was a notable shift in how the media portrayed the war, the U.S. government's role, and the South Vietnamese government. Reporters became increasingly aware of the complexities and controversies surrounding U.S. involvement, and the media's perspective started to change, showing more skepticism and criticism. Notably, The Tet Offensive in 1968 marked a shift towards more skeptical press coverage. This event is where the media begins to break with the

government over Vietnam. It had become obvious by this time that the government was not telling the whole truth at the very least, and the media had gotten sick of it (Vaughan, 2020). Following that, the release of the 1971 Pentagon Papers was a critical moment, reinforcing the media's skepticism about the U.S. government's transparency and intentions in Vietnam. Recognizing these events, how did the New York Times coverage transform regarding topic, tone, and other linguistic patterns from 1967 to 1973, following the timeline that captures the pre- and post-Pentagon Papers eras?

Natural Language Processing (NLP), the computational analysis of human language, effectively processes and interprets large text datasets. It incorporates methods like Topic Modeling and Sentiment Analysis, extensively utilized in historical newspaper analysis for uncovering evolving themes and sentiments. Sentiment Analysis, a key NLP technique, discerns the probable tone or emotion in text by assigning numerical scores to words and phrases. A notable example is UC Berkeley's project *Parsing the news: Using NLP to analyze news articles*. This initiative scrutinized the sentiment (positive, negative, neutral) in 400,000 quotations from approximately 77,000 U.S. news articles about nuclear weapons, spanning 2011-2022. UC Berkeley utilized the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool for this sentiment analysis. VADER is especially tailored for media texts, such as those from The New York Times, accommodating unique aspects of media language, including slang, emojis, and abbreviations. VADER, easily integrable with Python, excels in not just evaluating words independently but also in considering their contextual environment on sentiment, such as punctuation, capitalization, and conjunctions. Another case study is the *Sentiment Analysis on New York Times Articles Data* project. Here, Carvajal et al. gathered 21,457 articles on Latino Immigration published between 1981 and 2020 from the New York Times, utilizing Python and

the newspaper's public API with specific keyword and time parameters. This study also compared VADER with another sentiment analysis tool, TextBlob, demonstrating VADER's superior efficiency in dictionary-level analysis and categorizing pronouns and spellings based on frequency. These studies highlight the possibility of extracting and analyzing large datasets of over 20,000 news articles for sentiments and linguistic patterns, showcasing the power and versatility of NLP in understanding and interpreting media narratives.

Methodology

This project uses various methods to analyze The New York Times' coverage of the Vietnam War from 1967 to 1973. The process begins with the extraction of a comprehensive dataset of 31,416 headlines. Utilizing The New York Times API and Python-based web scraping techniques, we will efficiently transform these headlines into a computer-readable format.

Since the New York Times API does not allow us to extract full articles, our analysis will focus on the headlines. Following data collection, Python and OpenRefine will be used for data cleaning and preprocessing. This includes removing irrelevant content (such as stopwords), handling missing values, and standardizing text formats. This step is crucial for segmenting the headlines into analyzable units.

To address the first research question, I analyze sentiment using the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool from the NLTK library. VADER is particularly well-suited for analyzing sentiment in social media text but also performs well with headlines and news articles due to its focus on the polarity (positive/negative) and intensity (strength) of sentiments expressed.

Each headline will be assigned a compound sentiment score ranging from -1 (most negative) to +1 (most positive). For ease of interpretation, these scores will be rescaled to a 1 to

10 scale. The rescaled sentiment scores will then be plotted over defined periods to visualize trends.

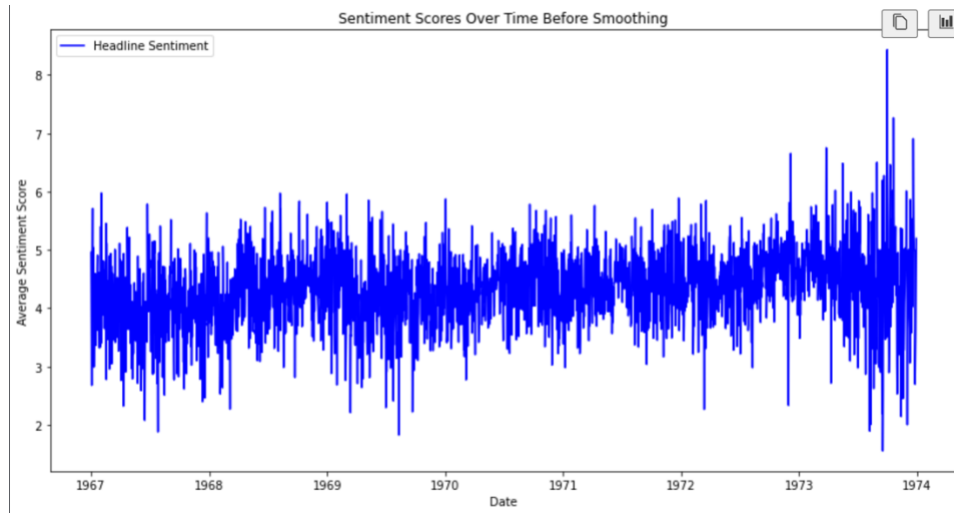


Figure 1: The sentiment scores of New York Times headlines over time before smoothing

A Gaussian filter will be applied to smooth out fluctuations and better observe these trends. This filter, commonly used for noise reduction and signal smoothing, applies a weighted average to neighboring data points, emphasizing the central value.

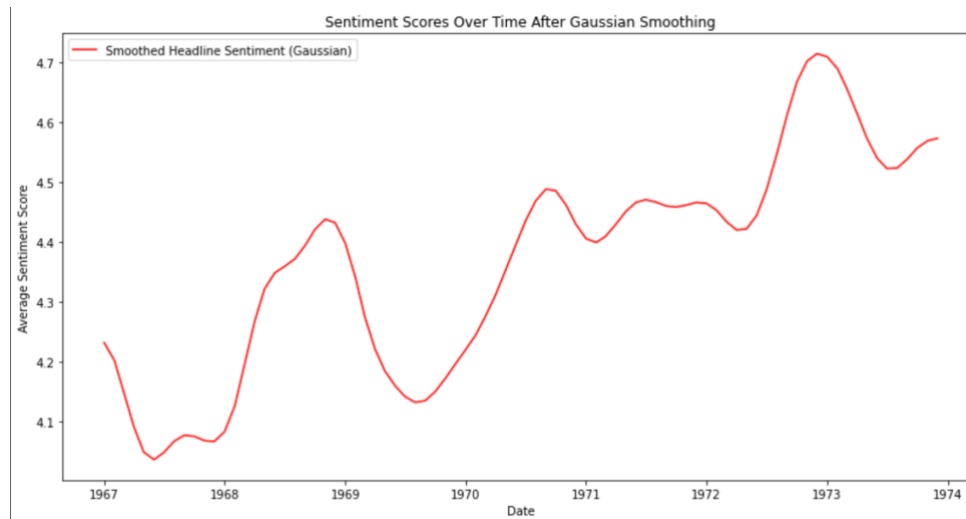


Figure 2: The sentiment scores of New York Times headlines over time after smoothing

For the second question, we will generate word clouds for positive and negative headlines. Positive headlines have high sentiment scores, while negative headlines have low sentiment scores.

The word clouds will help visually extract and compare the most frequent words and phrases used in each category, highlighting the differences in predominant themes associated with political figures, geographical locations, and the human cost of the war between positive and negative headlines.

Hence, we can discern periods of both optimism and pessimism, providing insight into how the media's portrayal of the Vietnam War evolved and how it reflected and possibly influenced public opinion. Finally, the final product is an accessible website developed using Shorthand. Shorthand was chosen for its user-friendly interface and ability to create visually compelling, narrative-driven content. The website ensures accessibility for all users, adhering to web content accessibility guidelines (WCAG).

Results

A. Sentiment Analysis

In our graphs, the x-axis represents the time period (monthly or yearly), marking the dates over which the sentiment scores of the headlines were analyzed. Otherwise, the y-axis represents the average sentiment score of the headlines, which has been smoothed using a Gaussian filter. The sentiment scores likely range from a scale where higher values indicate more positive sentiment.

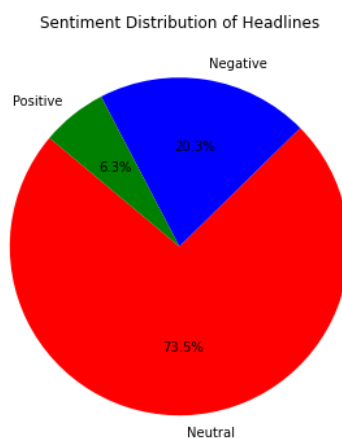


Figure 3: Sentiment Distribution of Headlines

Most headlines during this period were neutral, accounting for 73.5% of the coverage. This high percentage reflects the journalistic norm of delivering factual news reports focused on essential details such as places, dates, and events without explicitly expressing a positive or negative tone. The neutrality in headlines indicates an attempt to provide unbiased information, allowing readers to form their own opinions based on the reported facts.

A significant portion of the headlines conveyed a negative sentiment, comprising 20.3% of the total. This negativity can be attributed to the numerous challenges and grim realities associated with the Vietnam War. Factors such as increasing American casualties, controversial military strategies, and political turmoil contributed to the negative tone. Additionally, the release of the Pentagon Papers in 1971 and the subsequent public outrage likely played a role in the prevalence of negative headlines. Positive headlines were the least frequent, making up only 6.3% of the total. These instances of positive sentiment were typically associated with moments of hope or progress, such as successful peace talks or temporary ceasefires. However, such instances were relatively rare compared to the overall negative and neutral coverage, highlighting the predominance of challenging and distressing news during this period.

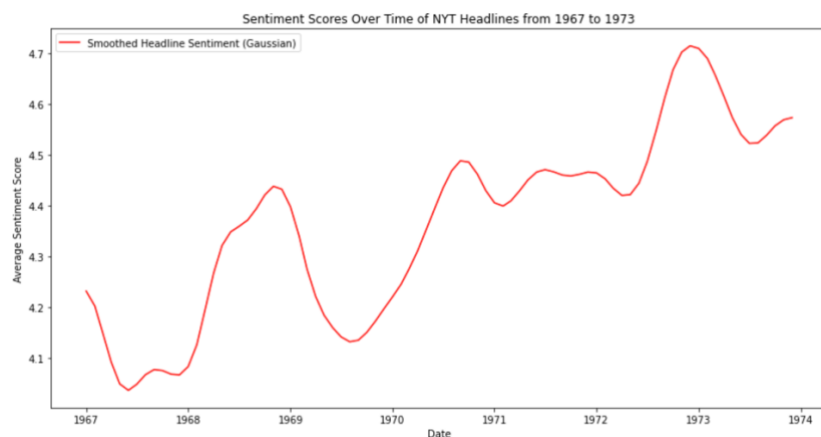


Figure 4: Sentiment Scores of New York Times headlines from 1967 to 1973

Overall, the sentiment score of New York Times headlines about Vietnam War from 1967 to 1973 has significant fluctuations in public sentiment over this critical period. In early 1967, sentiment scores in The New York Times' coverage of the Vietnam War begin relatively high but show a steep decline, reaching their lowest point in 1968. This period coincides with significant escalations in the Vietnam War, most notably the Tet Offensive in early 1968. The Tet Offensive, a coordinated series of North Vietnamese attacks on more than 100 cities and outposts in South Vietnam, marked a major turning point in public perception of the war. The negative sentiment during this time likely reflects the increasing disillusionment and frustration with the war's progress and the growing number of casualties. Following the low sentiment scores in 1968, there is a sharp rise in sentiment scores through 1969, suggesting a temporary improvement in public mood. This period coincides with the beginning of peace talks, which started in May 1968, and the anticipation of potential resolutions to the conflict. The initiation of peace negotiations likely provided a sense of hope and optimism, contributing to the positive shift in sentiment.

Between 1969 and 1971, sentiment scores show a generally upward trend with some fluctuations. This period witnessed significant developments, including President Nixon's policy of Vietnamization, which aimed to reduce American involvement by shifting combat roles to South Vietnamese forces. The initial optimism surrounding these policies might have contributed to the improving sentiment. However, the release of the Pentagon Papers in 1971 was a pivotal moment that dramatically altered public sentiment and media coverage. The documents exposed government deceit about the war, leading to a sharp decline in trust and an increase in negative sentiment. The New York Times, as one of the key publishers of the Pentagon Papers, played a crucial role in disseminating this information, influencing the sentiment scores reflected in their articles.

In the period from 1971 to 1973, sentiment scores reach a peak around 1972 before experiencing another decline. This peak may correspond to the signing of the Paris Peace Accords in January 1973, which was perceived as a hopeful step toward ending the war. The Accords aimed to establish peace in Vietnam and end the war, leading to a temporary boost in public sentiment. However, the subsequent decline reflects the lingering uncertainties and challenges in the aftermath of the accords, as the reality of achieving lasting peace and stability in Vietnam remained elusive.

B. Word Clouds

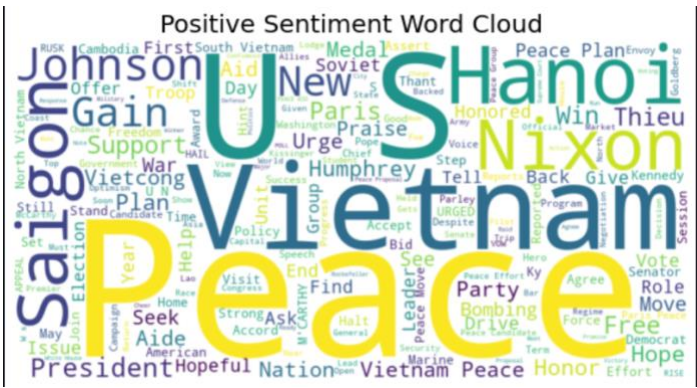


Figure 5: Word Cloud of Positive-sentiment Headlines

The positive sentiment word cloud from New York Times headlines about the Vietnam War between 1967 and 1973 highlights a strong focus on peace efforts and diplomatic actions. Prominent words such as “Peace,” “Hope,” “Praise,” “Accord,” and “Honor” suggest a significant emphasis on resolving the conflict and achieving peace. Key political figures like “Nixon,” “Johnson,” “Thieu,” and “Humphrey,” as well as critical locations such as “Saigon,” “Hanoi,” and “Paris,” feature prominently, indicating their involvement in peace initiatives and negotiations. The presence of supportive terms like “Support,” “Aid,” “Help,” “Plan,” “Accept,” and “Join” reflects the positive actions and collaborative efforts taken by leaders and nations to bring an end

executed following a structured methodology that included a comprehensive literature review, data collection and analysis, prototype development, and rigorous evaluation.

The final product, [an accessible website](#) created using Shorthand, demonstrates the practical application of our research findings. The website not only presents the research in an engaging and visually compelling manner but also adheres to web content accessibility guidelines, ensuring that it is usable by a broad audience.

Conclusions

The analysis of New York Times headlines from 1967 to 1973 provides insight into the shifting public sentiment and media portrayal of the Vietnam War, particularly around the pivotal year of 1971. The sentiment scores reveal a dynamic landscape influenced by key events such as the release of the Pentagon Papers and ongoing peace negotiations.

Prior to the release of the Pentagon Papers, sentiment scores showed a general upward trend, reflecting a period of cautious optimism. This optimism was likely driven by ongoing peace negotiations and President Nixon's policy of Vietnamization, which aimed to gradually withdraw U.S. troops and transfer combat responsibilities to South Vietnamese forces.

Following the publication of the Pentagon Papers, sentiment scores dipped significantly. The media's coverage captured the initial shock and outrage over the government's deception, leading to a more negative portrayal of the war and its management. This period marked a significant shift in public sentiment as trust in the government eroded.

The release of the Pentagon Papers intensified media scrutiny of the government's actions and policies related to the Vietnam War. Articles became more critical, mirroring the public's growing cynicism and distrust. Despite the initial dip, sentiment scores peaked around 1972, likely due to the ongoing peace talks and the eventual signing of the Paris Peace Accords in January

1973. However, the subsequent decline in sentiment reflects the lasting effects of the Pentagon Papers, as public trust in the government continued to diminish.

Also, in the earlier years, headlines predominantly highlighted the intense combat, casualties, and military operations, as evidenced by frequent references to words like “War,” “Enemy,” “Attack,” and “Killed.” This focus underscores the media’s attention to the immediate and violent aspects of the conflict, portraying a war-torn Vietnam and the significant human cost involved. Simultaneously, peace initiatives were covered substantially, particularly as the war progressed. Prominent terms such as “Peace,” “Negotiations,” “Accords,” and references to the “Paris Peace Accords” indicate an increasing media emphasis on diplomatic efforts to resolve the conflict. The presence of political figures like Presidents Nixon and Johnson, as well as key locations like Saigon, Hanoi, and Paris, reflects the geopolitical significance and the central role of the United States in both the conflict and peace processes.

The word clouds also suggest a narrative shift over time, from a primary focus on the horrors and immediacy of war towards a more balanced view that includes substantial coverage of peace efforts and political maneuvers. This shift highlights the evolving media portrayal of the Vietnam War, moving from the battlefield to the negotiating table.

In conclusion, the New York Times headlines from this period encapsulate the multifaceted nature of the Vietnam War, illustrating a journey from relentless conflict to persistent yet challenging peace efforts. This analysis underscores the media's significant role in shaping public perception and understanding of historical events through its coverage of both the brutal realities of war and the hopeful aspirations for peace.

Discussion

We encountered limitations with our main dataset that constrained our exploration. The New York Times API's limitation to headlines and lead paragraphs posed a significant challenge, as the full articles could provide a richer dataset for analysis. While we explored methods to scrape full articles, the sheer volume of 31,416 articles proved unmanageable. Consequently, we decided to focus our analysis on headlines, acknowledging the limitations this imposed on the depth and accuracy of our sentiment analysis.

Moreover, headlines inherently capture only the main points and essential information of articles, potentially omitting the complete tone and purpose conveyed in the full articles. Important nuances and detailed analyses present in the body of the articles might be lost, leading to an incomplete understanding of the overall sentiment. Thus, the sentiment expressed in headlines may not fully represent the depth and complexity of the reporting.

The sentiment analysis of headlines does not account for the increasingly skeptical and pessimistic tone of reporting during this era. Walter Cronkite's 1968 assessment that the conflict was "mired in stalemate" marked a significant shift in media reporting, reflecting growing disillusionment with the war. This shift may have mirrored rather than influenced public sentiment, highlighting the complex interplay between media portrayal and public perception. Also, reporting from Vietnam was largely uncensored, with only a few instances involving violations of military security by journalists. This freedom allowed for a diverse range of perspectives and stories. Despite this, dominant narratives often emphasized high casualty rates and the challenges faced by American forces, contributing to public disillusionment with the war. The uncensored nature of the reporting ensured a wide array of viewpoints but also highlighted the harsh realities of the conflict.

While the media played a crucial role in shaping public opinion, the increasing American casualties were a primary factor in undermining support for the war. As casualties rose, public support waned, indicating that the sentiment expressed in headlines was just one of many elements influencing public perception. The broader context of the war and its impact on American society played a significant role in shaping the public's views and reactions.

Future Recommendations

One significant avenue for future research is the expansion of the dataset to include entire articles rather than just headlines and lead paragraphs. This expansion would enable a more comprehensive content analysis, allowing researchers to capture the full scope of the authors' tone, intent, and rhetorical strategies. Full articles provide richer context and a deeper understanding of the narrative structures and linguistic nuances used by journalists. By analyzing complete texts, researchers can better assess the overall sentiment and thematic development within news stories.

Additionally, incorporating articles from other major newspapers, such as The Washington Post, The Los Angeles Times, and international sources, would provide a more diverse perspective. This broader dataset would help mitigate publication-specific biases and offer a more balanced view of media coverage. By comparing and contrasting different sources, researchers can identify patterns and variations in reporting styles and biases, contributing to a more nuanced understanding of media landscapes.

Another recommendation is to explore advanced Natural Language Processing (NLP) techniques to handle language complexity and rhetorical elements more effectively. Techniques such as sentiment-aware embeddings and transformer-based models (e.g., BERT, GPT) can provide a more nuanced understanding of the sentiments expressed in the articles. These models are adept at capturing subtleties like sarcasm, irony, and rhetorical questions, which are often

present in sophisticated journalistic writing. Discourse analysis methods can also be employed to analyze the structure and flow of arguments within articles, further enriching the sentiment analysis.

Furthermore, developing a multi-dimensional sentiment framework could significantly enhance the analysis. This framework would go beyond the traditional positive, negative, and neutral sentiment categories to include dimensions such as trust, anticipation, fear, and surprise. By mapping articles onto a broader sentiment spectrum, researchers can gain a more detailed understanding of the emotional and psychological impact of media reports on the public. Such a framework would allow for a more granular analysis of how different emotions are conveyed and perceived in journalistic content, providing deeper insights into media influence and public sentiment.

Bibliography

- Becker, Elizabeth. "The Secrets and Lies of the Vietnam War, Exposed in One Epic Document." *The New York Times*, The New York Times, 9 June 2021, www.nytimes.com/2021/06/09/us/pentagon-papers-vietnam-war.html.
- Bollinger, Lee C., 1946, Geoffrey R. Stone, and Ohio Library and Information Network. *National Security, Leaks, and Freedom of the Press: The Pentagon Papers Fifty Years on*. Edited by Lee Bollinger, and Geoffrey Stone. Oxford University Press, New York, NY, 2021, doi:10.1093/oso/9780197519387.001.0001.
- Carvajal, Gabriel, Karl Schmitt, and Gregg B. Johnson. "Sentiment Analysis on New York Times Articles Data." Symposium on Undergraduate Research and Creative Expression (SOURCE), no. 917, 2020, <https://scholar.valpo.edu/cus/917>.
- Connelly, Matthew, et al. "New evidence and new methods for analyzing the Iranian Revolution as an intelligence failure." *Intelligence and National Security*, vol. 36, no. 6, 29 July 2021, pp. 781–806, <https://doi.org/10.1080/02684527.2021.1946959>.
- Grimmer, Justin, and Brandon M. Stewart. "Text as data: The promise and pitfalls of Automatic Content Analysis Methods for political texts." *Political Analysis*, vol. 21, no. 3, 2013, pp. 267–297, <https://doi.org/10.1093/pan/mps028>.
- Hallin, Daniel C. *The "Uncensored War" the Media and Vietnam*. Univ. of California Press, 2010.

Kaiser, Jonas, Adrian Rauchfleisch, and Nikki Bourassa. "Connecting the (Far-)Right Dots: A Topic Modeling and Hyperlink Analysis of (Far-)Right Media Coverage during the US Elections 2016." *Digital Journalism*, vol. 8, no. 3, 2020, pp. 422-441.

I., Sanela. "Web Scraping the New York Times Articles with Python : Part i." *Medium*, CodeX, 22 Jan. 2023, medium.com/codex/web-scraping-the-new-york-times-articles-with-python-part-i-e2d6fc02d4e0.

Nguyen, Nam. "Combining Unsupervised Topic Modeling and Rule-Based Data Labeling to Analyze Research Articles On..." *Medium*, Dev Genius, 3 July 2023, blog.devgenius.io/analyzing-research-articles-on-vietnam-with-nlp-combining-unsupervised-topic-modeling-and-6b5b17edd6a6.

"Parsing the News: Using NLP to Analyze News Articles." *CDSS at UC Berkeley*, data.berkeley.edu/project/parsing-news-using-nlp-analyze-news-articles. Accessed 26 Jan. 2024.

Sheehan, Neil, E. W. Kenworthy 1909, and Fox Butterfield. *The Pentagon Papers: The Secret History of the Vietnam War*. Racehorse Publishing, New York, New York, 2017.

Spencer, Graham. *The Media and Peace: From Vietnam to the 'War on Terror'*. Palgrave Macmillan, New York; Basingstoke [England];, 2005.

Tomasi, Adam. "Topic Modeling Headline News About the Anti-Vietnam War Movement: A NULab Project," 14 May 2021, <https://cssh.northeastern.edu/nulab/topic-modeling-anti-vietnam-war-movement/>.

Vaughan, Brock J.. "War, Media, and Memory: American Television News Coverage of the Vietnam War." *Bridges: An Undergraduate Journal of Contemporary Connections*, vol. 4, no. 1, 2020, https://scholars.wlu.ca/bridges_contemporary_connections/vol4/iss1/5.

Wang, Guofeng, Yilin Liu, and Shengmeng Tu. "Discursive use of stability in New York Times' coverage of China: a sentiment analysis approach." *Humanities and Social Sciences Communications* 10.1 (2023): 1-10.

Yang, Tze-I., Andrew Torget, and Rada Mihalcea. "Topic modeling on historical newspapers." *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. 2011.